



Social learning promotes institutions for governing the commons

**Sigmund, K., De Silva, H., Hauert, C. and Traulsen,
A.**

**IIASA Interim Report
June 2010**



Sigmund, K., De Silva, H., Hauert, C. and Traulsen, A. (2010) Social learning promotes institutions for governing the commons. IIASA Interim Report. IR-10-053 Copyright © 2010 by the author(s). <http://pure.iiasa.ac.at/9432/>

Interim Report on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work. All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting repository@iiasa.ac.at



International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report

IR-10-053

Social learning promotes institutions for governing the commons

Karl Sigmund (ksigmund@iiasa.ac.at)
Hannelore De Silva (hannelore.desilva@wu.ac.at)
Christoph Hauert (christoph.hauert@math.ubc.ca)
Arne Traulsen (traulsen@evolbio.mpg.de)

Approved by

Ulf Dieckmann
Program Leader, EEP

June 2011

Interim Reports on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

Social learning promotes institutions for governing the Commons

Karl Sigmund^{1,2}, Hannelore De Silva³, Arne Traulsen⁴ & Christoph Hauert⁵

Theoretical and empirical research highlights the role of punishment in promoting collaborative efforts^{1,2,3,4,5}. However, both the emergence and the stability of costly punishment are problematic issues. How can punishers invade a society of defectors by social learning or natural selection, and how can second-order exploiters (who contribute to the joint effort but not to the sanctions) be prevented from drifting into a coercion-based regime and subverting cooperation? Here, we compare the prevailing model of peer-punishment^{6,7,8} with pool-punishment, which consists in committing resources, prior to the collaborative effort, to prepare sanctions against free-riders. Pool punishment facilitates the sanctioning of second-order exploiters, since these are exposed even if everyone contributes to the common good. In the absence of such second-order punishment, peer-punishers do better than pool-punishers, but with second-order punishment, the situation is reversed. Efficiency is traded for stability. Neither other-regarding tendencies or preferences for reciprocity and equity, nor group selection or prescriptions from higher authorities are necessary for the emergence and stability of rudimentary forms of sanctioning institutions regulating common pool resources and enforcing collaborative efforts.

Many economic experiments on 'public goods games' (PG games) have shown that a substantial fraction of players are willing to incur costs in order to impose fines on exploiters, i.e., those who do not contribute to the joint effort^{1,2,3,4,5,6,7,8}. As a consequence, the threat of punishment looms credibly enough to increase the average level of pro-social contributions. However, the sanctioning system is itself a public good. Thus punishers are often seen as altruistic, since others benefit from their costly efforts^{9,10,11,12,13}. Conversely, those who refrain from punishing exploiters are 'second-order free-riders'. Among self-interested agents, second-order free-riding should spread and ultimately cause the collapse of cooperation.

A solution is to also punish second-order free-riders¹⁴. But such 'second-order punishment' risks being subverted by third-order free-riders in turn, leading to infinite regress. Moreover, if everyone contributes to the public good, second-order free riders will not be spotted. Their number can grow through neutral drift, ultimately allowing defectors to invade with impunity.

We show how a simple mechanism can overcome this objection.

¹ Faculty of Mathematics, University of Vienna, A-1090 Vienna, Austria

² International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

³ WU (Vienna University of Economics and Business), A-1090 Vienna, Austria

⁴ Max Planck Institute for Evolutionary Biology, 24306 Ploen, Germany

⁵ Department of Mathematics, University of British Columbia, Vancouver V6T 1Z2, Canada

There exists a variety of sanctioning systems. Most experiments on PG with punishment have considered peer punishment: *after* the PG game, individuals can impose fines on exploiters, at a cost to themselves. Interestingly, the first experiment on PG with punishment¹⁵ considered a different mechanism. Here, players decide whether to contribute to a 'punishment pool' *before* contributing to the PG. This can be viewed as a first step towards an institutionalized mechanism for punishing exploiters, and compared with the self-financed contract enforcement games in Ostrom's 'Governing the Commons'¹⁶. It is like paying towards a police force, whereas peer-punishers take law-enforcement into their own hands.

Peer and pool punishment are both expensive ways to impose negative incentives on free-riders. In many economic experiments, the increase in cooperation is more than matched by the costs of punishment, and an overall *reduction* of total payoff is observed^{8,9}. Since the costs of pool punishment arise even when there are no exploiters to be punished, pool punishment seems even more socially expensive than peer punishment.

On the other hand, the issue of second order punishment favors pool punishment. If everyone contributes to the public good, then peer-punishers are not distinguishable from second-order free-riders. By contrast, pool-punishers declare themselves beforehand. We may expect that pool punishment leads more easily to a second-order punishment regime, and hence to more stability.

Since sanctioning institutions, as known from social history, usually forbid individuals to take the law into their own hands, it is also worthwhile to investigate the competition of peer vs. pool punishment.

A model based on evolutionary game theory shows that both peer and pool punishment can emerge, if participation in the joint effort is optional rather than compulsory. Pool punishment requires second-order punishment, whereas peer punishment is little affected by it. Both sanctioning mechanisms can evolve if players simply imitate whatever yields the highest payoff. If peer-punishers compete with pool-punishers, all depends on second-order punishment. Without it, the population is dominated by peer-punishers. With it, pool-punishers take over, although the average income is thereby reduced.

A 'punishment fund' can be viewed as a rudimentary institution to uphold the common interest. Many small-scale societies use this principle, for instance by hiring an enforcer. In 'Governing the Commons'¹⁶, several examples of self-financed contract enforcement are described. They concern the provisioning and the appropriation of common resources, for instance high mountain meadows (the proverbial 'Commons'), irrigation systems, or inshore fisheries. Our model shows that individuals can spontaneously adopt a self-governing institution to monitor contributions and sanction free riders. It needs no top-down prescriptions from higher authorities, nor great feats of planning: trial and error, and the imitation of successful examples, can lead to a social contract among individuals guided by self-interest.

To model a PG game, we assume that if $N \geq 2$ individuals participate in the interaction, each can decide whether to contribute a fixed amount $c > 0$ to the common pool. This amount will be multiplied by a factor $r > 1$ and then divided among the $N-1$ other players. If all contribute, they obtain $(r-1)c$ each. Since contributors do not benefit from their own contribution, self-interested players ought to contribute nothing. If all do this, their payoff will be 0. This reveals a social dilemma.

Pool-punishers not only contribute c to the PG game, but also, beforehand, an amount G to a punishment pool. Free-riders will be fined an amount BN_v proportional to the number N_v of pool-punishers. In the case of second-order punishment, second-order free-riders will also be fined the same amount. Peer-punishers contribute c to the PG game, and after the game impose a fine β on each free-rider in their group, at a cost γ . If N_w peer-punishers are in the group, each defector pays a total fine βN_w . In case of second-order punishment, second-order defectors are treated just like defectors.

Let us assume that the game is not compulsory^{11,17}. Some players may abstain from the joint enterprise. They can do something else instead, and earn a payoff σ independent of what others are doing. If only one player is willing to engage in the joint effort, there will be no PG game, and the solitary would-be participant earns σ like the non-participants.

Let M denote the population size, X the number of players who participate in PG games and contribute, but do not punish, Y that of defectors (who participate, but contribute neither to the PG game nor to the sanctions), Z that of non-participants, V that of pool- and W of peer-punishers. Random samples of N individuals are faced with the opportunity of a joint enterprise. Social learning leads to preferential copying of successful strategies. We obtain their long-run frequencies by numerical simulations (cf. Figs. 1, 2, 3). In a limiting case, we obtain analytic results (see SI) which we now describe.

Let us first neglect peer-punishment, and assume that the payoff σ for non-participants lies between 0 (obtained if all free-ride) and $(r-1)c-G$ (obtained if all contribute to the PG *and* the punishment pool). The inequality

$$0 < \sigma < (r-1)c - G \quad (1)$$

highlights that participating in the joint enterprise is a venture, which succeeds if most contribute, and fails if most don't.

In the absence of second-order punishment, the long-run frequencies in the (X,Y,Z,V) -population are $\frac{1}{7}(2,2,2,1)$, and little cooperation is achieved. With second-order punishment, the long-run frequencies are $(0,0,0,1)$. The population is dominated by pool-punishers enforcing cooperation. If the game is compulsory (i.e., no Z), the population consists of free-riders only.

Alternatively, if we neglect pool-punishment, and assume

$$0 < \sigma < (r-1)c, \quad (2)$$

the long-run frequencies in the (X,Y,Z,W) -population are $\frac{1}{M+8}(2,2,2,M+2)$ and punishers prevail, with or without second-order punishment. Again, if the game is compulsory, only free-riders subsist in the long run.

In the competition of peer- with pool-punishers *without* second-order punishment, peer-punishers win. The long-run frequencies in the (X,Y,Z,V,W) -population are $\frac{1}{3M+23}(6,6,4,1,3M+6)$. *With* second-order punishment, pool-punishers win, the frequencies are $(0,0,0,1,0)$.

Repression of free-riding is a basic theme for several major transitions in evolution¹⁸, and can lead to evolutionarily stable strategies allocating part of the contribution towards suppressing competition¹⁹.

In human societies, sanctions are ubiquitous^{4,16,20,21}. Peer punishment emerges more easily than pool punishment, since it requires no second-order punishment, and (2) is weaker than (1). But with second-order sanctions, pool punishment out-competes peer punishment, despite being socially expensive. Both types of punishment only emerge, in our model, if players can opt out of the joint enterprise. This restricts the range of applications^{22,23}. But there is considerable evidence that cooperation can increase, if participation is voluntary, rather than compulsory^{24,25,26} (see SI for an intuitive explanation).

Many early experiments on PG with punishment terminated after six or ten rounds, and while punishment usually raised the propensity to cooperate, the overall income was often less than without punishment^{2,8,9}. But if the number of rounds is sufficiently large, cooperation becomes common³. As long as players avoid antisocial punishment of contributors⁵ (a feature not included in our model), peer punishment becomes cost-free. Pool punishment entails fixed costs and thus is less efficient. On the other hand, peer punishment is ill-suited for second-order punishment, as has also been observed empirically²⁷. Pool punishment is more conducive to second-order punishment. A sanctioning institution should view anyone not contributing to its upkeep as defector, and resort to second-order punishment. Adding second-order punishment may add to the cost of sanctioning, but as long as (1) holds, results are unaffected.

Experimental PG games allowing players to opt, from round to round, between treatments with or without peer punishment²⁸, or to vote on whether to forbid antisocial punishment²⁹, suggest intermediary stages towards pool-punishment. Further steps towards endogenous institution formation are analyzed in^{23,30}. We considered players motivated entirely by self-interest, and did not assume preferences for reciprocity or equity²¹. This obviously does not mean that such preferences do not exist. Their emergence may actually have been favored by the prevalence of sanctioning institutions over thousands of years.

We left out many important issues, such as quorum-sensing and signaling, reputation and opportunism, repeated interactions, or graduated punishment, and did not specify how pool punishment is actually set up, or 'who guards the guardians'. Our model is minimalistic, but allows proof of principle. Origins of institutions are notoriously difficult to trace, but we have shown that they can emerge spontaneously among self-interested individuals.

Methods summary: We apply evolutionary game theory to populations of fixed size M and variable composition X, Y, Z, V, W (the frequencies of the five strategies for the optional PG game with peer or pool punishment). We compute the payoffs obtained by players using these strategies. The payoff differences define the probabilities that the strategies are copied through social learning, as a function of a parameter $s \geq 0$ measuring 'imitation strength'. Together with an 'exploration rate' $\mu \geq 0$ which specifies the propensity to switch randomly to another strategy, this defines a stochastic process describing the evolution of the frequencies X, Y, \dots . We compute their stationary distributions (which correspond to the relative frequencies in the long run) both numerically and, in a limiting case, analytically, and check these values by individual-based simulations. This allows to compare the evolution of any subset of the five strategies under social learning. For details on methods cf. SI.

References

1. Fehr, E. & Gächter, S., Cooperation and punishment in public good experiments, *American Economic Review* **90**, 980-994 (2000)

2. Rockenbach, B. & Milinski, M., The efficient interaction of indirect reciprocity and costly punishment, *Nature* **444**, 718-723 (2006)
3. Gächter, S., Renner, E. & Sefton, M., The Long-Run Benefits of Punishment, *Science* **322**, 1510-1512 (2008)
4. Henrich, J. *et al.*, Costly punishment across human societies, *Science* **312**, 176-177 (2006)
5. Herrmann, B., Thoeni, C. & Gächter, S., Antisocial punishment across societies, *Science* **319**, 1362-1367 (2008)
6. Fehr, E. & Gächter, S., Altruistic punishment in humans, *Nature* **415**, 137-140 (2002)
7. Gardner, A. & West, S.A., Cooperation and punishment, especially in humans, *Amer. Nat.* **164**, 753-764 (2004)
8. Egas, M. & Riedl, A., The economics of altruistic punishment and the maintenance of cooperation, *Proc. Roy. Soc. London B* **275**, 871-878 (2008)
9. Fehr, E. & Rockenbach, B., Detrimental effects of sanctions on human altruism, *Nature* **422**, 137-140 (2003)
10. Boyd, R., Gintis, H., Bowles, S., & Richerson, P., The evolution of altruistic punishment, *Proc. Natl. Acad. Sci. USA* **100**, 3531-3535 (2003)
11. Fowler, J.H., Altruistic punishment and the origin of cooperation, *Proc. Nat. Acad. Sci. USA* **102**, 7047--7049 (2005)
12. Nakamaru, M. & Iwasa, Y., The evolution of altruism and punishment: role of the selfish punisher, *J. Theor. Biol.* **240**, 475-488 (2006)
13. Lehmann, L., Rousset, F., Roze, D. & Keller, L., Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment, *American Naturalist* **170**, 21-36 (2007)
14. Boyd, R. & Richerson, P.J., Punishment allows the evolution of cooperation (or anything else) in sizable groups, *Ethology and Sociobiology* **3**, 171-195 (1992)
15. Yamagishi, T. The provision of a sanctioning system as a public good, *J. Pers. and Soc. Psych.* **51**, 110-116 (1986)
16. Ostrom, E., *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, (1990)
17. Hauert, C., Traulsen, A., Nowak, M.A., H. Brandt, H. & Sigmund, K., Via freedom to coercion: the emergence of costly punishment, *Science* **316**, 1905-1907 (2007)
18. Maynard Smith, J. & Szathmary, E., *The Major Transitions in Evolution*, Oxford UP, Oxford (1997)
19. Frank, S. A., Mutual policing and repression of competition in the evolution of cooperative groups, *Nature* **377**, 520-522 (1995)
20. Levin, S.A. (ed.), *Games, Groups, and the Global Good*, Springer New York (2009)

21. Falk, A., Fehr, E. & Fischbacher, U., Appropriating the commons: a theoretical explanation, in *The Drama of the Commons*, eds. Ostrom L et al. , National Academy Press, Washington DC, pp.157-191 (2002)
22. Mathew, S. & Boyd, R., When does optional participation allow the evolution of cooperation, *Proc. Roy. Soc. London B* **276**, 1167-1174 (2009)
23. Boyd, R., Gintis, H. & Bowles, S., Coordinated punishment of defectors sustains cooperation and can proliferate when rare, *Science* **328**, 617-620 (2010)
24. Orbell, J.H. & Dawes, R.M., Social welfare, cooperator's advantage, and the option of not playing the game, *Amer. Soc. Rev.* **58**, 787-800 (1993)
25. Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K., Volunteering as a Red Queen mechanism for cooperation, *Science* **296**, 1129-1132 (2002)
26. Semmann, D., Krambeck, H.J., & Milinski, M., Volunteering leads to rock-paper-scissors dynamics in a public goods game, *Nature* **425**, 390-393 (2003)
27. Kiyonari, T., Barclay, P., Wilson, M. & Daly, D., Second order punishment in one-shot prisoner's dilemma, *Int. Journ. Psych.* **39**, 329-334 (2004)
28. Gürer, O., Irlenbush, B., & Rockenbach, B., The competitive advantage of sanctioning institutions, *Science* **312**, 108--111 (2006)
29. Ertan, A., Page T., & Putterman, L., Who to punish? Individual decisions and majority rule in mitigating the free rider problem, *European Econ. Review* **53**, 495-511 (2009)
30. Kosfeld, M., Riedl A. & Okada A., Institution formation in public goods games, *American Econ. Review* **99**, 1335-1355 (2009)

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements: K.S. thanks TECT I-104 G15, A.T. the Emmy Noether program of the DFG, and C.H. NSERC (Canada).

Author Contribution: the authors contributed in equal measure.

Author Information: Correspondence should be addressed to K.S. (karl.sigmund@univie.ac.at). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions.

Figure legends:

Figure 1: Time-evolution of the competition of peer-punishment and pool-punishment. Shown are two typical individual-based simulation runs, without (**a**) and with (**b**) second-order punishment. In **a**, peer-punishers prevail most of the time, but sometimes second-order free riders invade. In this case, defectors and then non-participants take over before peer-punishment gets re-established. In **b**, pool-punishers eventually establish a very stable regime. Parameters: $N=5$, $r=3$, $c=1$, $\sigma=1$, $\gamma=\beta=0.7$.

$B=G=0.7$, $M=100$ and $\mu=10^{-3}$. The updating is by strong imitation ($s \rightarrow +\infty$), i.e., players with lower average payoff always imitate players with higher average payoff. The initial population consists of defectors only.

Figure 2: Pool punishment in compulsory and voluntary PG games. Results are shown for extensive simulations based on social learning (SI). In order to obtain reliable average frequencies, each player updates 10^7 times. Data points are supported by analytical approximations (solid lines). Parameters as in Fig.1, but with $\mu=10^{-6}$ and with variable imitation strength s . For small s updating is mostly random, and frequencies of all strategies are roughly equal. Discrimination between strategies increases with s . **(a)** compulsory PG games lead for larger s to a regime of defectors. **(b)** in voluntary PG games, the cycle $X \rightarrow Y \rightarrow Z \rightarrow X$ provides an escape out of the defectors' regime via recurrent opportunities for establishing a sanctioning system with second order punishment.

Figure 3: The competition of peer- and pool-punishers in voluntary PG games. **(a)** Without second-order punishment, peer-punishers prevail, but cooperation keeps breaking down and needs to be re-established (c.f. Fig.1a). **(b)** With second-order punishment, pool-punishers prevail. Simulations and parameters as in Fig.2.

Supplementary Information:
**Social learning promotes institutions for governing
the Commons**

Karl Sigmund ^{1,2}, Hannelore De Silva ³, Arne Traulsen ⁴ &
Christoph Hauert ⁵

¹ Faculty of Mathematics, University of Vienna, A-1090 Vienna, Austria

² International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

³ WU (Vienna University of Economics and Business), A-1090 Vienna, Austria

⁴ Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany

⁵ Department of Mathematics, University of British Columbia, Vancouver V6T 1Z2, Canada

May 16, 2010

Methods

The methods are based on a straightforward application of evolutionary game dynamics for finite populations. First, we discuss the dynamics based on social learning together with analytical approximations and implementations of individual based simulations. In section 1 we describe what happens in the absence of punishment, i.e. for three strategies: X -players participate, and contribute an amount c to the public goods (PG) game; Y -players participate, but do not contribute; and Z -players do not participate. With X , Y and Z , we also denote the number of players using the corresponding strategy (and $M = X + Y + Z$ is the total population size, which we assume to be constant). In sections 2 and 3, we additionally consider V -players, who contribute to the PG as well as to a punishment pool, with and without second-order punishment. In sections 4 and 5 this is repeated for W -players, who contribute to the PG game and then peer-punish (with or without second order punishment). Finally, in section 6, we address the competition of peer and pool punishment (i.e., $M = X + Y + Z + V + W$). In each section, we compute the average payoff values, and analyze a limiting case ('strong imitation').

Social learning

We assume that two players i and j are randomly chosen. Their expected payoff values P_i and P_j depend on the strategies of the two players and on the frequencies X, Y, \dots of the strategies. There are many ways to model social learning. We shall assume that player i adopts the strategy of player j with a probability which is an increasing function of the payoff difference $P_j - P_i$. A frequently used choice for this probability is

$$\frac{1}{1 + \exp[-s(P_j - P_i)]}, \quad (1)$$

where the 'imitation strength' $s \geq 0$ measures how strongly the players are basing their decisions on payoff comparisons^{31,32,33,34,35}. For $s \rightarrow 0$ (or for $P_i = P_j$), a coin toss decides whether to imitate or not. Small values of s correspond to a regime we call 'weak imitation'. In this case, imitation is basically random, but more successful players are imitated slightly more often. For $s \rightarrow +\infty$, i.e., 'strong imitation', a more successful player is always imitated, a less successful never. The homogeneous populations correspond to absorbing states of the stochastic process: once such a state is reached, imitation cannot produce any change. Thus we shall assume that additionally, with a certain probability $\mu > 0$ (the exploration rate), a player switches randomly to another strategy without imitating another player. The resulting Markov chain has a stationary distribution which, if the population size M is large and there are more than two strategies, requires considerable efforts to compute numerically. In addition to individual-based computer simulations, we shall consider the limiting case of very small exploration rates, the so-called 'adiabatic' case. In that case, if in a homogeneous population a single dissident arises, then its fate (elimination or fixation) will be settled through the imitation process before the next exploration step occurs.

More precisely, let us assume that there are d strategies $1, \dots, d$. By X_k we denote the number of players using strategy k ($\sum X_k = M$). The homogeneous population with $X_k = M$ will be denoted by All_k . With probability $\mu/(d-1)$, a single individual switches from k to $l \neq k$. The probability that subsequently, imitation leads to the fixation of the dissident strategy l is denoted by ρ_{kl} . The fixation probability can be computed by the formulas known from the

theory of birth-death processes^{34,35,36},

$$\rho_{kl} = \frac{1}{1 + \sum_{q=1}^{M-1} \prod_{X_l=1}^q \frac{T_{l \rightarrow k}(X_l)}{T_{k \rightarrow l}(X_l)}}.$$

In our case, the probability that one out of X_l players with strategy l is chosen as a focal player and imitates one of the $X_k = M - X_l$ players with strategy k is given by

$$T_{l \rightarrow k}(X_l) = \frac{X_l}{M} \frac{M - X_l}{M} \frac{1}{1 + \exp[-s(P_k - P_l)]},$$

where payoffs P_l and P_k depend on the number of l and k players, i.e., on X_l and $X_k = M - X_l$.

The fixation probability ρ_{kl} simplifies to

$$\rho_{kl} = \frac{1}{1 + \sum_{q=1}^{M-1} \exp[s \sum_{X_l=1}^q (P_k - P_l)]} \quad (2)$$

This form makes it easy to address the limit of strong imitation, $s \rightarrow +\infty$.

The probability of a transition from All_k to All_l is $\mu\rho_{kl}/(d-1)$. If the $d \times d$ transition matrix is mixing, it has a unique normalized left eigenvector to the eigenvalue 1, and this is the stationary distribution which describes the percentage of time (in the long run) spent by the state of the population in the vicinity of the homogeneous state All_k . One can show^{37,38} that the stationary distribution of the full system converges for $\mu \rightarrow 0$ to the stationary distribution of this 'embedded' Markov chain on the homogeneous states whose transition probabilities from All_k to All_l (for $k \neq l$) are given by $\rho_{kl}/(d-1)$ (μ cancels out). For numerical confirmation, we refer to Fig. 1 (in the Supplementary Information).

Simulations and numerical solutions

The individual based simulations mimic the social learning dynamics outlined above for arbitrary exploration rates, μ . Each individual achieves an average payoff based on random sampling of the interaction groups. This reflects a situation where individuals interact often and only occasionally reassess and update their strategies. With probability μ , players randomly adopt any other available strategy, and with probability $1 - \mu$, they update according to Eq. (1). The long-run mean frequency of each strategy is determined by averaging over $T > 10^7$ updates per player.

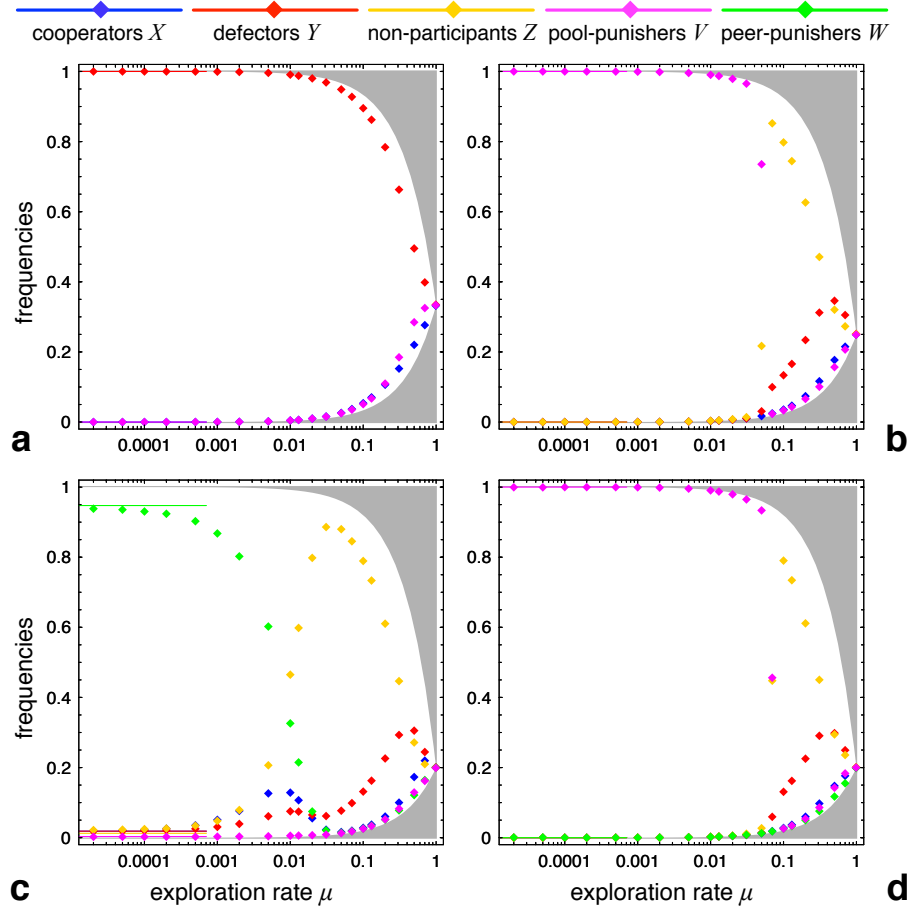


Figure 1: Scenarios of sanctioning in public goods games for variable exploration rates μ . For large μ , random exploration dominates, which results in roughly equal average frequencies of all available strategies. All strategies tend to be present in the population at all times. Because a fraction μ of the population always mutates, the minimum frequency of each strategy is μ/d (for d strategies) and the grey-shaded areas are inaccessible to the process. For smaller μ , the population spends increasing amounts of time in homogeneous states between subsequent mutations. (a) even though sufficiently large μ can push the population from the region of attraction of $AllY$ to $AllV$, the population is unable to remain near this cooperative state and defection dominates. (b) in voluntary public goods games, pool punishers prevail except for large μ , where risk-averse non-participants take over. (c, d) whether peer-punishers or pool-punishers prevail in voluntary public goods games depends on second-order punishment (c.f. Fig. 3). Without second-order punishment, peer-punishers prevail, c, but exploration rates μ of order 0.01 are large enough to destroy cooperation, so that non-participants prevail. With second-order punishment, pool-punishers dominate, (d), and since $AllV$ is strongly attracting, much larger μ -values are required before cooperation is destroyed. The data points, obtained by having each player update 10^7 times, are supported by analytical approximations (solid lines) for very small values of μ . Parameters: same as in Fig. 2 (main text), but with fixed imitation strength $s = 10$ and variable exploration rate μ .

Numerical computations of the stationary distribution (for small μ) based on the fixation probabilities in Eq. (2) and individual-based simulations show that the results hold not only for the limiting case, but for a large set of plausible values for the parameters $\mu, s, c, r, \gamma, \beta, M, G$ and B , see^{37,38}. For online experimentation, we refer to <http://www.hanneloredesilva.at/sanctions> and the *VirtualLabs* at <http://www.univie.ac.at/virtuallabs>.

1 No punishment

In a population consisting of X contributors and $Y = M - X$ defectors, random samples of N individuals play the PG game. A co-operator obtains on average

$$\begin{aligned} \sum_{k=0}^{N-1} \frac{\binom{X-1}{k} \binom{M-X}{N-1-k}}{\binom{M-1}{N-1}} \left(rc \frac{k}{N-1} - c \right) \\ = rc \frac{X-1}{M-1} - c \end{aligned}$$

(the summation variable k represents the number of other contributors, sampling is done without replacement, probabilities obey the hypergeometric distribution). Defectors obtain from the public good on average

$$\sum_{k=0}^{N-1} \frac{\binom{X}{k} \binom{M-1-X}{N-1-k}}{\binom{M-1}{N-1}} rc \frac{k}{N-1} = rc \frac{X}{M-1}.$$

Let us now assume that the population consists of X contributors, Y defectors and Z non-participants. The probability that the other $N-1$ players of a sample are unwilling to participate is

$$\frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}.$$

Hence the average payoff for defectors is

$$P_Y = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}} \sigma + \left(1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}} \right) rc \frac{M-Z-Y}{M-Z-1}, \quad (3)$$

that for contributors

$$P_X = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}} \sigma + \left(1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}} \right) c \left(r \frac{M-Z-Y-1}{M-Z-1} - 1 \right), \quad (4)$$

and of course $P_Z = \sigma$ (cf.³⁹). The three strategies form a Rock-Paper-Scissors cycle. More precisely, if $Z = 0$, defectors do always better than contributors ($P_Y > P_X$); but in the absence of contributors ($X = 0$), non-participants do better than defectors ($P_Z \geq P_Y$, with equality if and only if $Y = 1$); and in the absence of defectors ($Y = 0$), contributors do better than non-participants ($P_X \geq P_Z$, with equality if and only if $X = 1$).

The resulting stochastic process exhibits cycling behavior. It is clear that if most players use strategy X , then Y -players do better, and if most players use strategy Y , the Z -players do better. It is less obvious to see why, in a population where most players use Z , X players do best. But if most players are non-participants, PG groups are small. In that case, random fluctuations can lead to groups with mostly X -players, who do well, so that many imitate them. This relates to Simpson's paradox³⁹

For small exploration rates, the embedded Markov chain describing the transitions between $AllX$, $AllY$ and $AllZ$ is given by

$$\begin{pmatrix} 1 - \frac{1}{2}\rho_{XY} - \frac{1}{2}\rho_{XZ} & \frac{1}{2}\rho_{XY} & \frac{1}{2}\rho_{XZ} \\ \frac{1}{2}\rho_{YX} & 1 - \frac{1}{2}\rho_{YX} - \frac{1}{2}\rho_{YZ} & \frac{1}{2}\rho_{XY} \\ \frac{1}{2}\rho_{ZX} & \frac{1}{2}\rho_{ZY} & 1 - \frac{1}{2}\rho_{ZX} - \frac{1}{2}\rho_{ZY} \end{pmatrix}. \quad (5)$$

The normalized left eigenvector to the eigenvalue 1 gives the stationary distribution, which by Eq. (2) can be evaluated numerically as a function of the imitation strength s . This is the basis of the analytical approximation under weak mutation shown in Figs. 2 and 3 (main text) and Fig. 1 (SI). In the limiting case of strong imitation, $s \rightarrow +\infty$, the embedded Markov chain simplifies significantly; for instance, $\rho_{XZ} = 0$ and $\rho_{XY} = 1$. Interestingly, $\rho_{ZX} = 1/2$. The reason is that a single X -mutant in a population of Z -players cannot participate in any game, and has payoff σ like the other non-participants. The next change obtained through the imitation process is equally likely to revert the X -player to the fold or to produce a second X -player. From then on, X -players have an expected payoff larger than σ and will increase to fixation. The transition matrix between $AllX$, $AllY$ and $AllZ$ reduces for $s \rightarrow \infty$ to

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix} \quad (6)$$

and the stationary distribution (the left eigenvector to the eigenvalue 1) is given by $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. The same argument used to simplify the dynamics for small μ and $s \rightarrow \infty$ is used in the discussion about punishment below.

2 Pool Punishment

Let us now assume in addition that V of the M players engage in pool punishment. This means that when N_v of them find themselves in a PG game, they not only contribute c to the public good, but pay an extra fee G towards the punishment pool. The fine of each exploiter will be proportional to the number of punishers, and hence of the form $N_v B$, for some $B > 0$.

For the moment, we neglect the possibility of second-order punishment (i.e. the punishment of non-punishers). The payoffs for non-participants and contributors are therefore unaffected. The payoff for pool-punishers satisfies

$$P_V = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}} \sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}) [c(r \frac{M-Z-Y-1}{M-Z-1} - 1) - G]. \quad (7)$$

Indeed, punishers only pay a fee into the pool if another player is willing to participate. The payoff for defectors is

$$P_Y = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}} \sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}) cr \frac{M-Z-Y}{M-Z-1} - \frac{B(N-1)V}{M-1}. \quad (8)$$

(If there is at least one punisher among the $N-1$ co-players in the sample, the PG game is played.)

If we assume that a population of pool-punishers does better than the non-participants, i.e., that

$$\sigma < (r-1)c - G, \quad (9)$$

then we obtain, in the case of strong imitation, the following embedded Markov chain for the transitions between the states $AllX$, $AllY$, $AllZ$ and $AllV$:

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{pmatrix}. \quad (10)$$

The explanation for the $\frac{1}{6}$ terms is the same as that for the $\frac{1}{4}$ in (6). If, in an *AllZ*-population, a mutation produces a single *X*-player, this player finds no partners for the PG game and obtains the same payoff as the non-participants. The next change obtained through the imitation process is equally likely to revert the *X*-player to the fold or to produce a second *X*-player. From then on, *X*-players have an expected payoff larger than σ and will increase to fixation.

The unique stationary distribution is given by $\frac{1}{7}(2, 2, 2, 1)$. This corresponds to two rock-paper-scissors cycles, one from *AllY* to *AllZ* to *AllX* and back to *AllY* again, the other (four-membered) from *AllY* to *AllZ* to *AllV* to *AllX* and back to *AllY*. Computer simulations confirm that the four homogeneous states supersede each other.

If the game is compulsory, i.e., if there are no *Z*-players, then the transitions between the states *AllX*, *AllY* and *AllV* are given by

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \quad (11)$$

and the stationary distribution is $(0, 1, 0)$. Free-riders take over.

3 Second-order pool-punishers

Let us now assume that the second-order exploiters, i.e., the *X*-players, are also punished. Thus their payoff is given by

$$P_X = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\sigma + \left(1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\right)c\left(r\frac{M-Z-Y-1}{M-Z-1} - 1\right) - \frac{B(N-1)V}{M-1}. \quad (12)$$

The other payoff values remain unchanged.

If pool-punishers can invade non-participants, i.e., (9) holds, the embedded Markov chain is given by in the case of strong imitation by

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (13)$$

The unique stationary distribution is $(0, 0, 0, 1)$, which means that punishers prevail.

4 Peer Punishment

Let us now assume instead that W players in the population engage in peer punishment. Each peer-punisher imposes a fine β on each defector in his or her sample, at a cost γ . Thus if there are N_y defectors and N_w peer-punishers in the sample, each defector pays a total fine $N_w\beta$, and each punisher incurs a cost $N_y\gamma$. We first omit second-order punishment.

In the absence of pool punishment, i.e. if $M = X + Y + Z + W$, the average payoff for punishers is

$$P_W = P_X - \frac{(N-1)Y}{M-1}\gamma \quad (14)$$

where P_X is given by (4), and the defectors' payoff is given by the expression in (3), reduced by

$$\frac{(N-1)W}{M-1}\beta. \quad (15)$$

For strong imitation, the embedded Markov chain on the states $AllX$, $AllY$, $AllZ$ and $AllW$ has the transition matrix

$$\begin{pmatrix} \frac{2}{3} - \frac{1}{3M} & \frac{1}{3} & 0 & \frac{1}{3M} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{3M} & 0 & 0 & 1 - \frac{1}{3M} \end{pmatrix}. \quad (16)$$

If, for instance, W -dissidents arise in an X -population, they do as well as the residents (all contribute, no one punishes), and the fixation probability in this 'neutral case' is $1/M$. It is easy to see that this Markov chain has a unique stationary distribution, given by

$$\frac{1}{M+8}(2, 2, 2, M+2). \quad (17)$$

For instance, if the populations size is $M = 92$, then for 94 percent of the time, the population is dominated by peer punishers.

5 Second order peer-punishers

Let us now assume that peer-punishers engage in second-order punishment: thus they impose fines β on the contributors too, at a cost γ to themselves.

If $M = X + Y + Z + W$, the average payoff P_X for contributors is given by (4), reduced by the average fine

$$\frac{(N-1)W}{M-1}\beta\left(1 - \frac{\binom{M-Y-2}{N-2}}{\binom{M-2}{N-2}}\right) \quad (18)$$

and the peer-punishers' payoff by (14), reduced by the average cost

$$\frac{(N-1)X}{M-1}\gamma\left(1 - \frac{\binom{M-Y-2}{N-2}}{\binom{M-2}{N-2}}\right) \quad (19)$$

for meting out extra punishment. The term $(1 - \dots)$ corresponds to having at least one defector in the sample (otherwise a punisher cannot be aware that the contributor does not punish).

In the limiting case of strong imitation, the Markov chain is exactly as before. Indeed, during the imitation process the population never consists of more than two types. Hence second-order punishment (which requires that W -players see that X -players fail to punish Y -players) will never occur.

6 The competition of pool- and peer-punishers

The outcome is: without second order punishment, pool-punishers lose and peer-punishers predominate in the long run. With second-order punishment, it is just the reverse. (We assume that pool-punishers punish peer-punisher, since these do not contribute to the punishment pool. It seems less likely that peer-punishers will punish pool-punishers, and we shall not assume it here. However, we stress that this assumption does not really matter. The reason: in a population with peer- and pool-punishers only, peer-punishment is not used and the pool-punishers do not reveal that they do not engage in it.)

Let us first consider the case without second-order punishment. For strong imitation, the embedded Markov chain describing the transitions matrix between $AllX$, $AllY$, $AllZ$, $AllV$

and $AllW$ is

$$\begin{pmatrix} \frac{3}{4} - \frac{1}{4M} & \frac{1}{4} & 0 & 0 & \frac{1}{4M} \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4M} & 0 & 0 & 0 & 1 - \frac{1}{4M} \end{pmatrix}. \quad (20)$$

The unique stationary distribution is $\frac{1}{3M+23}(6, 6, 4, 1, 3M + 6)$. This means that the majority consists of peer-punishers. In the case with second-order punishment, the matrix is

$$\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{4M} & 0 & 0 & 0 & 1 - \frac{1}{4M} \end{pmatrix}. \quad (21)$$

In this case, the stationary distribution is $(0, 0, 0, 1, 0)$. Pool-punishers win.

References

31. McFadden, D. Structural Discrete Probability Models Derived from Theories of Choice, in *Structural Analysis of Discrete Data and Econometric Applications*, Manski C.F. & McFadden D.L. (Eds), Cambridge, MIT Press (1981)
32. Blume, L.E. The statistical mechanics of strategic interaction *Games Econ. Behav.* **5**, 387-424 (1993)
33. Szabó, G. & Toke, C. Evolutionary prisoner's dilemma game on a square lattice, *Phys. Rev. E* **58**, 69-73 (1998)
34. Traulsen, A., Nowak, M.A. & Pacheco, J.M. Stochastic dynamics of invasion and fixation, *Phys. Rev. E* **74**, 011909 (2006)
35. Nowak, M.A. *Evolutionary Dynamics*, Cambridge MA, Harvard UP (2006)

36. Karlin, S., & Taylor, H.E. *A First Course in Stochastic Processes*, 2nd ed., Academic Press, London
37. Fudenberg, D. & Imhof, L. Imitation Processes with Small Mutations, *Journal of Economic Theory* **131**, 251-262 (2005)
38. Antal, T. & Scheuring, I. Fixation of strategies for an evolutionary game in finite populations, *Bull. Math. Biol.* **68**, 1923- (2006)
39. De Silva, H., Hauert, C., Traulsen, A. & Sigmund, K. Freedom, enforcement, and the social dilemma of strong altruism, *J. Evolutionary Economics* **20** 203-217 (2010)